



Morphological Classification of Infrared Galaxies Based on WISE

Zhi-Ren Pan¹, Bo Qiu¹, Cui-Xiang Liu¹, A-Li Luo², Xia Jiang¹, and Xiao-Yu Guo¹

¹ School of Electronic and Information Engineering, Hebei University of Technology, Tianjin 300401, China; qiubo@hebut.edu.cn

² Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China

Received 2023 September 21; revised 2024 January 3; accepted 2024 February 13; published 2024 April 17

Abstract

This study introduces a novel convolutional neural network, the WISE Galaxy Classification Network (WGC), for classifying spiral and elliptical galaxies using Wide-field Infrared Survey Explorer (WISE) images. WGC attains an accuracy of 89.03%, surpassing the combined use of K-means or SVM with the Color–Color method in more accurately identifying galaxy morphologies. The enhanced variant, WGC_mag, integrates magnitude parameters with image features, further boosting the accuracy to 89.89%. The research also delves into the criteria for galaxy classification, discovering that WGC primarily categorizes dust-rich images as elliptical galaxies, corresponding to their lower star formation rates, and classifies less dusty images as spiral galaxies. The paper explores the consistency and complementarity of WISE infrared images with SDSS optical images in galaxy morphology classification. The SDSS Galaxy Classification Network (SGC), trained on SDSS images, achieved an accuracy of 94.64%. The accuracy reached 99.30% when predictions from SGC and WGC were consistent. Leveraging the complementarity of features in WISE and SDSS images, a novel variant of a classifier, namely the Multi-band Galaxy Morphology Integrated Classifier, has been developed. This classifier elevates the overall prediction accuracy to 95.39%. Lastly, the versatility of WGC was validated in other data sets. On the HyperLEDA data set, the distinction between elliptical galaxies and Sc, Scd and Sd spiral galaxies was most pronounced, achieving an accuracy of 90%, surpassing the classification results of the Galaxy Zoo 2 labeled WISE data set. This research not only demonstrates the effectiveness of WISE images in galaxy morphology classification but also represents an attempt to integrate multi-band astronomical data to enhance understanding of galaxy structures and evolution.

Key words: methods: data analysis – infrared: galaxies – galaxies: spiral – galaxies: elliptical and lenticular – cD

1. Introduction

The infrared band is very important for astronomical observations. Infrared astronomical telescopes have captured massive image data and their performance has constantly been improved, in such cases as the Infrared Astronomical Satellite (IRAS) (Neugebauer et al. 1984), Infrared Space Observatory (ISO) (Metcalf & Kessler 1990), Spitzer Space Telescope (Spitzer) (Werner et al. 2004), Wide-field Infrared Survey Explorer (WISE) (Wright et al. 2010) and James Webb Space Telescope (JWST) (Gardner et al. 2006) that launched in 2021 December. Infrared wavelengths can reveal many details that cannot appear in visible light, such as the dust and faint celestial objects, which makes the study of infrared images valuable.

WISE data are used in this paper. From 2009 to 2011, WISE surveyed the entire sky and produced a wealth of infrared data. WISE is composed of four infrared bands—W1, W2, W3 and W4 with wavelengths of 3.4, 4.6, 12 and 22 microns, respectively (Wright et al. 2010). The sensitivity of WISE was much deeper than any other large-scale infrared explorer when it was launched.

Celestial bodies display varying shapes when observed through different astronomical telescopes. For instance, the Sloan

Digital Sky Survey (SDSS) captures detailed features of galaxies, such as their morphology and spiral arms, as Ahumada et al. (2020) described. Consistent with the Hubble classification scheme, galaxies are categorized into spiral, elliptical, lenticular and irregular types. However, due to the relatively lower resolution of WISE, its images lack distinct spiral arm features, primarily revealing color characteristics. Projects like Galaxy Zoo, a citizen science initiative, need help accurately classifying data, which even professional astronomers find difficult. This can lead to biases in final category voting. Although researchers at Galaxy Zoo endeavor to mitigate these biases using methods like weighting and thresholding, some bias is inevitably present (Domínguez Sánchez et al. 2018). The distinct characteristics of WISE infrared images compared to SDSS optical images result in different tendencies in classifier predictions trained on these data sets. When the predictions from both data sets are consistent, galaxy morphology categories can be identified with higher accuracy. A clean galaxy data set aids astronomers in analyzing different types of galaxies. The differing features between the two data sets may also imply complementarity in galaxy morphology classification tasks, suggesting that integrating both types of data could result in a classifier with higher accuracy.

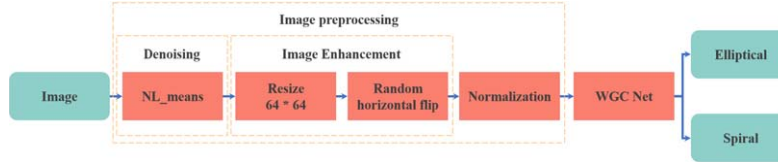


Figure 1. Process flow chart of the infrared image classification system. The workflow begins with WISE image preprocessing, which is first subjected to denoising to mitigate noise interference. Subsequently, image enhancement techniques such as resizing and random flipping are applied to bolster the robustness of the model. The refined images are then input into WGC, which classifies the galaxies as spiral or elliptical.

As proposed by Wright et al. (2010), the Color–Color diagram of WISE has demonstrated the effectiveness of W1–W2 and W2–W3 features in distinguishing between elliptical and spiral galaxies. However, there is a certain overlap in the feature space between these two types of galaxies, making it challenging to separate them completely. In light of this, our study explores the possibility of using images for the morphological classification of galaxies. With the advancement of convolutional neural networks (CNN), extracting more effective features from a large set of images has become feasible. Researchers like Dieleman et al. (2015), Mittal et al. (2020) and Zhang et al. (2022) have conducted galaxy classification studies based on SDSS images and CNNs. In the context of WISE infrared data, machine learning methods have been extensively applied to tasks such as redshift estimation and active galactic nucleus detection, for instance, the photometric redshift estimation of QSOs by Curran (2020) and Kunsági-Máté et al. (2022), as well as Faisst et al. (2019)’s work in active galactic nucleus detection. In the morphological classification within WISE, Zhao et al. (2023) already performed a trinary classification of stars, galaxies and quasars using WISE images. In this paper, we have designed a CNN model specifically for classifying WISE galaxy images, with the system’s workflow diagram illustrated in Figure 1.

The contributions are as follows: (1) The introduction of the WISE Galaxy Classification Network (WGC), a network based on WISE images, and WGC_mag, which integrates WISE images and magnitude data for classifying galaxies using WISE data. (2) Exploring the consistency and complementarity between WISE infrared images and SDSS images in galaxy morphology classification tasks. When predictions from WGC and the SDSS Galaxy Classification Network (SGC) concur, galaxy categories can be identified with an accuracy of 99.30%. The Multi-band Galaxy Morphology Integrated Classifier (MGMIC) was proposed to capitalize on the complementarity of WISE and SDSS images, utilizing their combined features. MGMIC achieves an accuracy of 95.39%, surpassing the 94.64% accuracy of SGC.

The structure of this paper is as follows: The data source and image synthesis method of WISE are introduced in Section 2. The WGC network and WGC_mag network are introduced in Section 3. The experimental results are presented in Section 4. In Section 5, the K-means and support vector machine (SVM) algorithms are used to classify galaxies based on the Color–

Color diagram. The experimental results are analyzed in Section 6. Section 7 explores the consistency and complementarity between WISE infrared images and SDSS images in galaxy morphology classification tasks. Section 8 evaluates the versatility of the classifier by applying WGC on the HyperLEDA data set. The paper concludes in Section 9.

2. Data Acquisition and Image Synthesis and Preprocessing

2.1. Data Acquisition

First, the labels, R.A. and decl. of galaxies are obtained from Galaxy Zoo 2 (Willett et al. 2013). Then, W1, W2 and W3 band data are cross-matched from WISE³ based on R.A. and decl. In order to prevent too much background noise from affecting the image classification accuracy, all the images’ size are set to $3 \times 34 \times 34$ (channels * height * width), approximately $50''$. The Color–Color diagram of WISE (Wright et al. 2010) uses W1, W2 and W3 bands to describe astronomical objects, so we also use them to synthesize and classify the images. There are two other reasons for using only W1, W2 and W3 band data, besides referencing Color–Color. Its angular resolution of $12''0$ is markedly higher than the $6''\text{--}6''5$ resolutions of W1, W2 and W3. Additionally, the sensitivity of the W4 band is noticeably lower than that of W1, W2 and W3 (Wright et al. 2010). These discrepancies result in the conspicuous reduction of morphological details in celestial bodies observed in the W4 band compared to those observed in W1, W2 and W3.

Galaxy Zoo 2 includes 11 tasks and 37 responses, with help from volunteers to identify images of galaxies and determine whether the images are spiral or elliptical, or not galaxies. We finally get 12,852 elliptical galaxies and 12,088 spiral galaxies from Galaxy Zoo 2. The parameters are displayed in Table 1.

2.2. Image Synthesis

The training of deep learning requires a lot of image data. Use of the `make_lupton_rgb` function from the Astropy package (Lupton et al. 2004) is widespread among researchers for synthesizing SDSS RGB images (González et al. 2018;

³ The data are available from WISE: <https://irsa.ipac.caltech.edu/applications/wise/>.

Table 1
Galaxy Parameter Filtering

Class	Task	Galaxy Zoo 2 Threshold Setting	Amount	Redshift
Spiral	T01	$f_{\text{features}/\text{disk}} \geq 0.430$	12088	0–0.0736
	T02	$f_{\text{edge-on,no}} \geq 0.715$		
	T04	$f_{\text{spiral,yes}} \geq 0.619$		
Elliptical	T01	$f_{\text{smooth}} \geq 0.469$	12852	0–0.0828
	T07	$f_{\text{completely_round}} \geq 0.5$		

Note. $f_{\text{features}/\text{disk}}$ is the frequency of smooth and disk-like structure, $f_{\text{edge-on,no}}$ is the frequency of an image without lateral edges and $f_{\text{spiral,yes}}$ is the frequency of an image that is a spiral galaxy.

Burke et al. 2019; He et al. 2021; Shi et al. 2022), as seen in the SDSS images of Galaxy Zoo 2 (Willett et al. 2013). Following these approaches, the `make_lupton_rgb` function is adopted for the composition of WISE RGB images. Before applying this function, it is essential to normalize the W1, W2 and W3 bands, which vary in their flux magnitudes. For this study, the W3 band is designated for red, W2 for green and W1 for blue. The brightness factor Q of `make_lupton_rgb` is set at 2, and the stretch factor at 0.5. Table 2 shows the synthesized images. This configuration effectively highlights the galaxy centers while distinctly revealing the dust around them.

2.3. Image Preprocessing

Images undergo a series of preprocessing steps before being input into the network, as shown in Figure 1, which primarily includes denoising and image enhancement. At the end of the image preprocessing, the image is normalized.

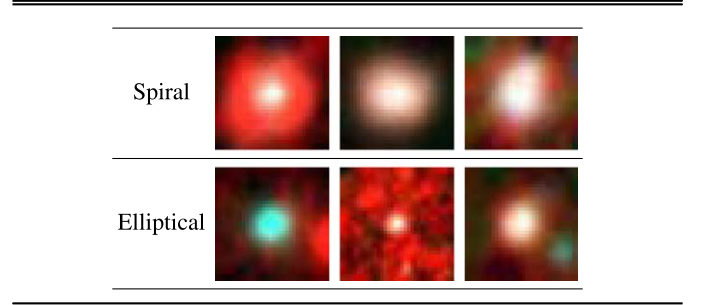
The denoising method used in this paper is Non-Local Means (NL-means) (Buades et al. 2005). The NL-means denoising algorithm effectively removes noise by comparing similar regions across the entire image and calculating a weighted average while preserving important details and structures within the image. As demonstrated in Figure 2(a), the denoised image is smoother than the original image. Due to the small sizes of WISE images, up-sampling the image to 64×64 is necessary. In order to improve the robustness, the operation of flipping the image horizontally with a probability of 0.5 is also chosen. The effectiveness of enhancement is shown in Figure 2(b).

To investigate whether image preprocessing has an effect in the experimental results, Comparative Experiment 1 is done and the result is shown in Table 6.

3. Design of WGC Network

3.1. Reasons for Choosing CNN

The most important part in image classification tasks is feature extraction. Traditional image feature extraction

Table 2
Examples of Synthesized Samples Using `make_lupton_rgb` Function

algorithms mainly include Histogram of Oriented Gradients (HOG) (Dalal & Triggs 2005), Local Binary Patterns (LBP) (Ojala et al. 2002), Haar (Viola & Jones 2001), Scale-Invariant Feature Transform (SIFT) (Lowe 2004), etc. The disadvantage of these algorithms is that suitable features need to be manually selected for different data sets; the more algorithms that are selected, the larger the feature dimension obtained. For infrared images with few texture details and small targets, the above algorithms cannot well represent the essential characteristics of the image.

In recent years, CNNs have been widely used in astronomical image classification, and the combination of CNNs and astronomical images has accelerated the development and research of astronomy. The basic idea of CNN is to build a function by building a multi-layer network. By increasing the depth and width of convolution layers and nonlinear mapping of activation functions, a CNN can greatly express abstract features of the data. CNN-based classification network models are becoming more and more popular and the classification accuracy is gradually improving.

At present, a variety of neural network classification models have been developed, such as the VGG series (Simonyan & Zisserman 2014), the ResNet series (He et al. 2015) and the EfficientNet series (Tan & Le 2019). These classification models and their improved versions have been successfully applied to various public data sets, including tasks related to galaxy classification. For instance, Zhu et al. (2019) and Gupta et al. (2022) conducted galaxy morphological classification of SDSS photometric images using an improved version of ResNet, each surpassing 90% accuracy in their respective data sets. Furthermore, Kalvankar et al. (2020) explored galaxy morphological classification using EfficientNet, attaining a high accuracy of 93.7% and an F1-score of 88.57% in a task involving seven distinct categories of galaxies.

3.2. WGC Network

The WGC network (as shown in Figure 3) is an end-to-end model. It is mainly composed of five basic convolution modules, four fully connected modules and two convolutional

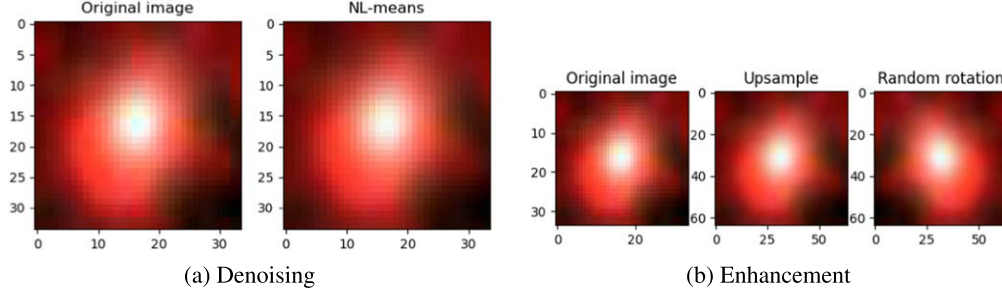


Figure 2. Example of image preprocessing. (a) illustrates the denoising operation, where NL-means is applied to remove noise and smooth the image; (b) depicts the image enhancement operations, including upscaling pixels to improve resolution and random rotations to enhance classifier robustness.

block attention modules. The high-level features of an image can be extracted by five basic convolution modules. The size of the convolution kernel in the five basic convolution modules decreases in turn. A large convolution kernel has a large receptive field so that more image information can be obtained after convolution, which is suitable for the WISE images with the characteristics of small image size and large target.

In each convolutional layer, we introduce a batch normalization (BN) layer in order to prevent gradients from disappearing or exploding and to speed up the training. It also solves the problem of covariate shift that occurs during the training process.

The activation function of the hidden layer is ReLU (Glorot et al. 2011). ReLU has three advantages. (1). ReLU can reduce computation. (2). ReLU successfully solves the gradient dispersion problem of the sigmoid activation function when the network is deep. (3). ReLU alleviates the overfitting problem.

The Convolutional Block Attention Mechanism (CBAM) (Woo et al. 2018) is added to the convolution module. Figure 4 shows the structure of CBAM. CBAM is a simple and effective attention module for a feedforward CNN. It is mainly composed of channel attention and spatial attention modules. CBAM is a lightweight network that does not add a lot of parameters to the network, and can also be trained end-to-end with the WGC network. CBAM enhances the model's ability to extract features and works well for networks with uncomplicated structures.

A fully connected layer is used to classify images. In the fully connected module, the number of neurons in the second and third fully connected layers is 2048. Too many neurons will lead to overfitting. Therefore, the dropout module is added after these two layers to remove some neurons with a probability of 0.5.

3.3. WGC_mag Network

Multimodal feature fusion is an important approach in the field of pattern recognition. Feature fusion methods are able to combine multiple modal features to enable neural networks to make judgements from multiple perspectives and reach the most correct conclusions.

Based on the WGC network, the WGC_mag network is designed, which incorporates image features and magnitude parametric features. As illustrated in Figure 5, the image is convolved and flattened into a one-dimensional vector with 256 elements. The parameters W1, W2 and W3, which are magnitudes corresponding to each image, are stitched onto this vector to obtain a one-dimensional vector with 259 elements. Finally the spliced vector is fed into the fully connected module.

4. Experiment

4.1. Evaluation Indicators

The experimental results are analyzed by a confusion matrix. The schematic diagram of the confusion matrix is shown in Table 3.

Accuracy is the proportion of all correctly judged results to the total observations; precision is the proportion that the model predicts correctly among all the results predicted by the model; recall rate is the proportion that the model predicts correctly among all the results for which the true value is positive; F1-score is the harmonic mean of precision and recall. The larger the F1-score is, the higher the quality of the model. The calculation formulas for accuracy, precision, recall and F1-score are shown in Equations (1), (2), (3) and (4) respectively:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (1)$$

$$\text{precision} = \frac{TP}{TP + FP}, \quad (2)$$

$$\text{recall} = \frac{TP}{TP + FN}, \quad (3)$$

$$F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}. \quad (4)$$

The values floating point operations (Flops), params, training time and total memory are also used to evaluate the model. Flops is used to calculate the amount of floating point operations and is a standard for considering the amount of computation in a network model. Params is used to represent the total number of parameters that need to be trained in the network model. Time taken to train the model for 150 epochs is

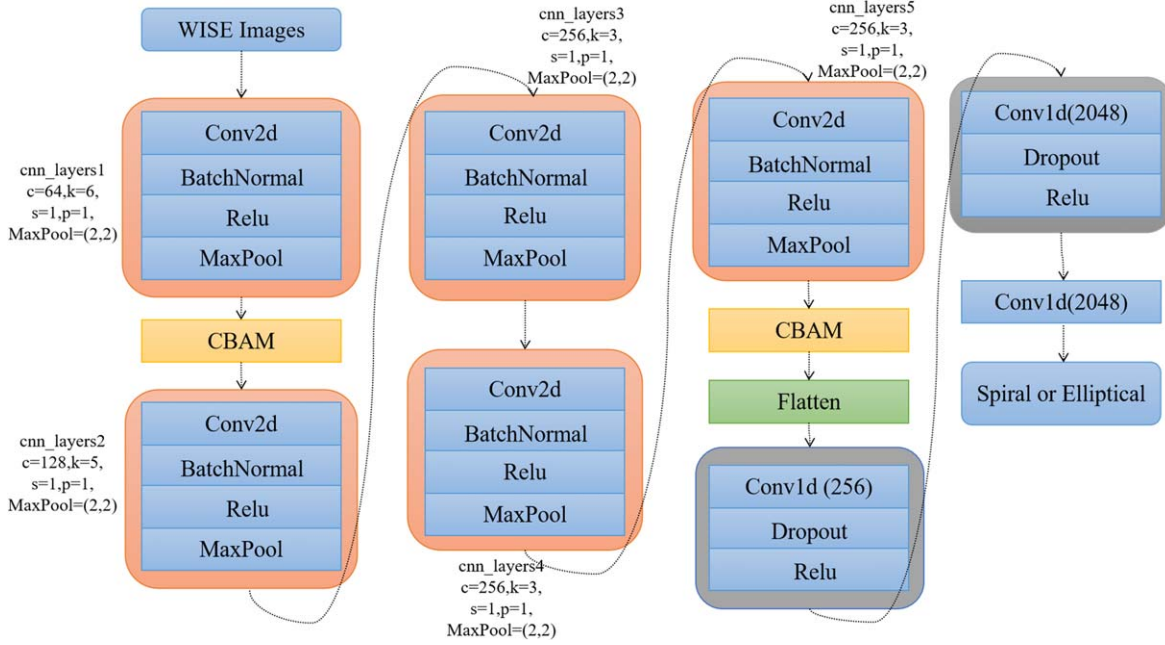


Figure 3. Diagram of WGC network: c is the number of channels, k is the size of the convolution kernel, s is the step size of the convolution, p is the padding size, Maxpool is the largest pooling layer and the last fully connected layer is marked as the number of neurons.

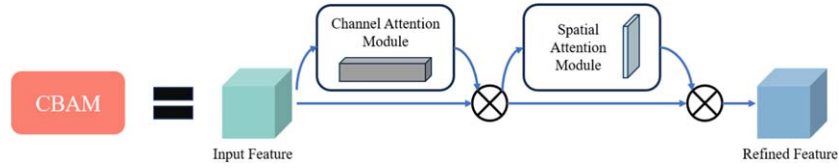


Figure 4. Structure of the CBAM. CBAM integrates attention mechanisms along both channel and spatial dimensions.

named the training time. Total memory is the memory occupied by training only one image.

4.2. Data Set Partitioning

All programs in this paper are Python programs, running on a desktop with i7-11700KF @ 3.60 GHz CPU, 32.0 GB memory and 64-bit Windows system, and accelerated with NVIDIA GeForce RTX 3070 Ti GPU. In the model training process, since the batch size depends on the size of the data sets and the processing power of the GPU, we set the batch size to 256. A total of 150 epochs were trained.

In this paper, the data sets are divided into training set, validation set and test set according to the ratio of 7:2:1. The specific data information is shown in Table 4. All images are preprocessed.

4.3. Experimental Results and Verification

The confusion matrix of the WGC network is shown in Table 5.

Comparative Experiment 1: Verifying the effectiveness of image preprocessing.

Image denoising can remove image details. In order to verify whether there is an impact on the experimental results, this paper analyzes the two types of experimental results based on the test set, as shown in Table 6 (Raw data mean without denoising and enhancement). After preprocessing, the accuracy is increased by 0.47%, the recall of the spiral galaxy is increased by 0.02 and the precision of the elliptical galaxy is increased by 0.01. Image preprocessing not only does not remove the image details but also improves the accuracy.

Comparative Experiment 2: Ablation experiment of CBAM.

Experiment 2 is an ablation experiment of CBAM. Ablation experiments refer to understanding the network by removing parts of the network and studying the performance of the network. Figure 6 shows the accuracy change graph of the training set and the validation set. Observing the training change curves of WGC_No_CBAM and WGC, the accuracy curve of the WGC network rises relatively stably but the WGC_No_CBAM model is particularly unstable. Comparing the experimental results of the test set, the accuracy increased by 0.55% after adding the CBAM. The comparison results above show that the existence of CBAM makes the whole

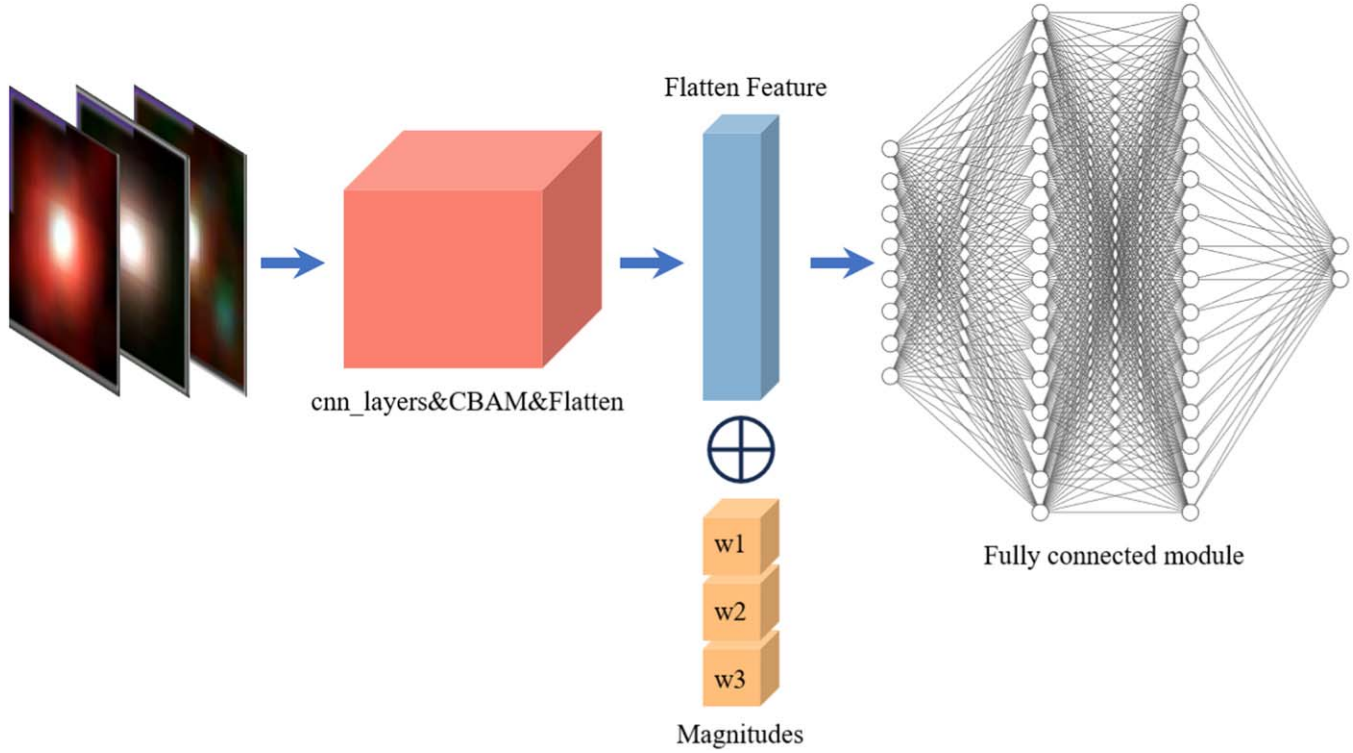


Figure 5. WGC_mag network. In line with the concept of multimodal feature fusion, the magnitude information from W1 to W3 is concatenated to the end of the flattened feature extracted by WGC. A fully connected layer processes the fused multimodal information to perform the classification operation.

Table 3
Schematic Diagram of Confusion Matrix

		Predicted value	
		True	False
Actual value	True	TP	FN
	False	FP	TN

Note. FN: False Negative, positive case is judged as a negative case. FP: False Positive, negative case is judged as a positive case. TN: True Negative, negative case is judged as a negative case. TP: True Positive, positive case is judged as positive case.

model converge faster and pay more attention to the useful information.

Comparative Experiment 3: Verifying the Effectiveness of WGC.

The WGC network is compared with VGG19 (Simonyan & Zisserman 2014), ResNet50 (He et al. 2015), EfficientNet_b0 (Tan & Le 2019), EfficientNet_b3 (Tan & Le 2019) and EfficientNet_v2 (Tan & Le 2021). The core of ResNet is the residual structure. By using the residual structure, ResNet makes it possible to train networks with hundreds of layers, resulting in a very powerful characterization capability. ResNet50 contains a total of 49 convolution layers and a

Table 4
Dataset Information

Class	Training Set	Validation Set	Test Set	Total
Spiral	8464	2408	1216	12088
Elliptical	8964	2568	1320	12852
Total	17428	4976	2536	24940

Table 5
Confusion Matrix of WGC Network

	Spiral	Elliptical
Spiral	1106	110
Elliptical	168	1152

neuron number of 2048 fully connected layers. The VGG19 model contains 16 convolution layers and 3 fully connected layers. In 2019, the performance of EfficientNet on the ImageNet validation set was demonstrated to be stronger than other networks (Tan & Le 2019). In 2021, EfficientNet_v2 became the strongest classification network. This paper calls the model that has been encapsulated in Python, and a classification layer with two neurons is added at the end.

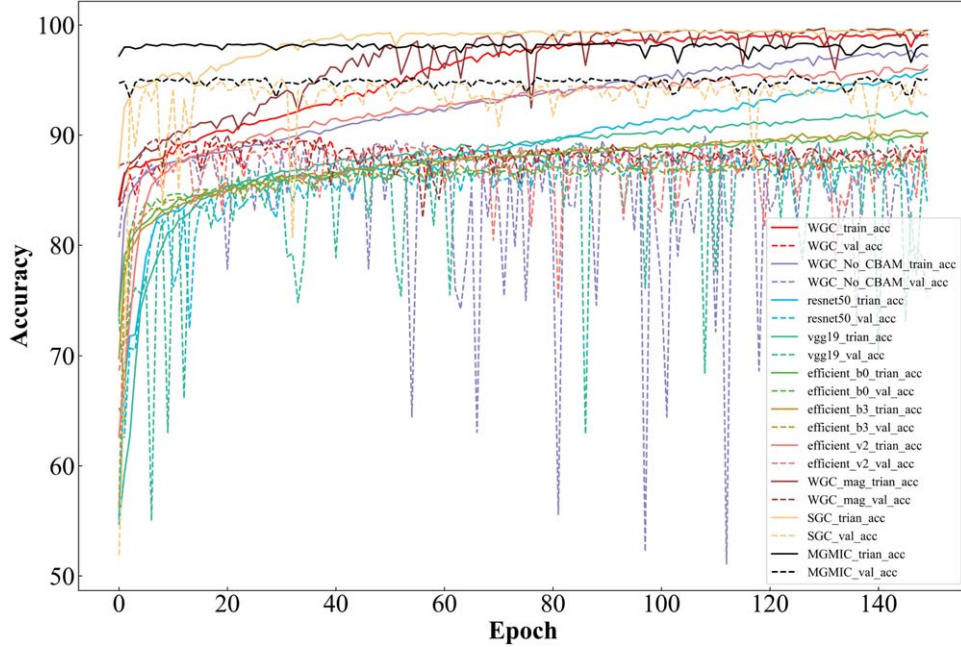


Figure 6. Accuracy of different model training and validation sets. WGC_No_CBAM is the WGC network without CBAM. The epoch is 150. The line “—” represents the accuracy curve of the training set, and the line “- -” represents the accuracy curve of the validation set.

Table 6
Results of Comparative Experiment 1. Bold Entries in the Table Highlight the Best Results in each Column

	Accuracy	Spiral Precision	Spiral Recall	Elliptical F1-score	Elliptical Precision	Elliptical Recall	Elliptical F1-score
Raw data	88.56%	0.89	0.87	0.88	0.88	0.90	0.89
WGC	89.03%	0.87	0.91	0.89	0.91	0.87	0.89

The accuracies of the training set and validation set of six CNN models are shown in Figure 6. Comparing WGC with other models, within the range of 60 epochs, the accuracy of the training set of the WGC network increases the fastest. In the validation set, the accuracy curve of the WGC network is particularly flat after 30 epochs, while ResNet50 has some obvious fluctuations, and VGG19 is unstable. Comparing the test set evaluation data (Table 10), the accuracy of the WGC network is the highest.

As shown in Table 11, WGC has an obvious advantage in parameters Training time and Total memory. WGC has the shortest training time and uses the least amount of memory when batch_size = 1, which means WGC can train more images in the shortest time compared to other networks.

Comparative Experiment 4: Verifying the Effectiveness of WGC in Classifying Galaxy Images with Single versus Multiple Astronomical Objects.

This experiment validated the impact of other sources’ contaminants on the classification of galaxies in WISE images. The existing data set was divided into two categories:

“Multiple” and “Single.” “Multiple” refers to images containing other celestial bodies in addition to the target galaxy, while “Single” refers to images that only have the target galaxy; Table 7 displays sample images. The TOPCAT software tool (Taylor 2005) was used to perform cross-matching within the image range to obtain the corresponding star catalog and determine whether multiple targets are present in the images. Table 8 presents the distribution of the two types of data. Overall, the ratio of “Single” to “Multiple” is approximately 4:5.

Independent tests were performed on each subset of the data, with the corresponding results detailed in Table 9. A comparative analysis reveals that for the “Multiple” data set, the accuracy of the WGC classifier experiences a marginal decline of 1.31% compared to the “Single” data set. This finding suggests that the presence of extraneous sources within the images exerts a measurable, yet relatively minor, impact on the performance of the WISE galaxy classifier.

Comparative Experiment 5: Compare WGC with WGC_mag.

Table 7
Examples of Single/Multiple Images in WISE

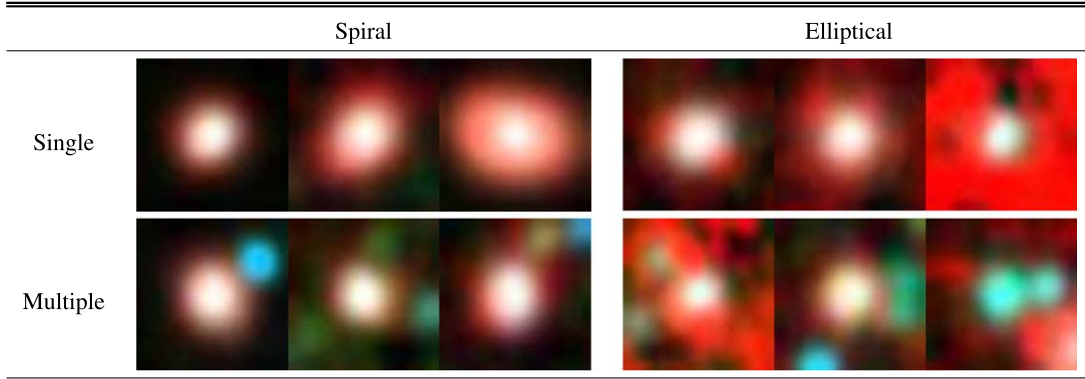


Table 8
Distribution of “Single” and “Multiple” Types of Data in the Dataset

	Train (Single:Multiple)	Validation (Single:Multiple)	Test (Single:Multiple)	Total (Single:Multiple)
Elliptical	3733:5230	1097:1471	613:707	5443:7408
Spiral	3940:4521	1094:1314	579:637	5593:6472

Table 9
Classification Performance of WGC on Single/Multiple Test Set. Bold Entries in the Table Highlight the Best Results in each Column

	Accuracy	Spiral Precision	Spiral Recall	Elliptical F1-score	Elliptical Precision	Elliptical Recall	Elliptical F1-score
Single	89.26%	0.89	0.88	0.89	0.88	0.90	0.89
Multiple	87.95%	0.88	0.87	0.87	0.87	0.90	0.88
Total	88.56%	0.89	0.87	0.88	0.88	0.90	0.89

As shown in Table 10, the WGC_mag network increased the accuracy of classification by nearly 1% by fusing image and magnitude features. The precision, recall and F1_score have all improved for different types of galaxies. However, as not all galaxies have both image and magnitude data, there are significant limitations to the use of WGC_mag.

5. Classifying Galaxies Based on the Color–Color Diagram

To demonstrate that images are more effective than magnitude values for galaxy classification tasks, traditional machine learning algorithms like K-means and SVM, combined with magnitude values, are utilized for classifying galaxies. Figure 7 displays the density distributions for two types of galaxies, W2–W3 and W1–W2. The most significant

difference between these galaxy types is observed in the W2–W3 distribution. The range for elliptical galaxies in W2–W3 is from 0.5 to 2, whereas it is from 2.5 to 4.5 for spiral galaxies. Furthermore, there is considerable overlap between the distributions, but the high-density areas have almost no overlap, similar to the Color–Color diagram distributions by Wright et al. (2010). Since only two feature values characterize each galaxy, there is no need for complex algorithms, leading to the choice of the commonly used K-means algorithm and SVM. SVM is a supervised learning algorithm primarily used for classification tasks, which finds an optimal hyperplane that maximizes the margin between different classes in the feature space. K-means is an unsupervised learning algorithm used for clustering, which partitions data into k distinct clusters by minimizing the variance within each cluster and maximizing the difference between clusters.

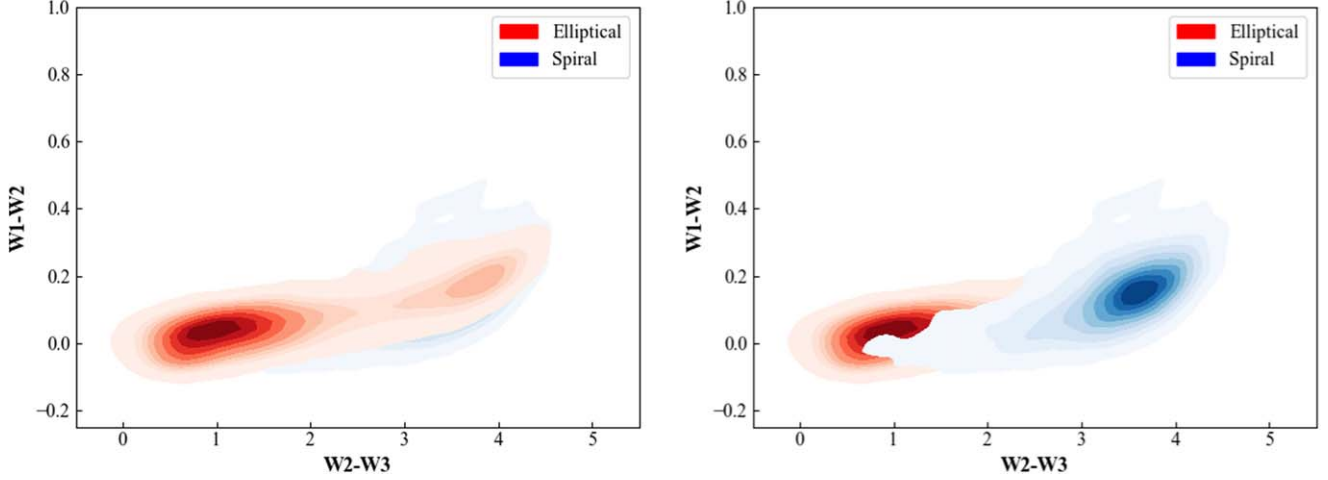


Figure 7. Color–Color diagram of test set. The left figure and right figure clearly show the distribution of elliptical and spiral galaxies. The darker the color is, the higher the density.

Table 10
Comparison of Experimental Results for Different Networks (1). Bold Entries in the Table Highlight the Best Results in each Column

Data Type	Network	Accuracy	Spiral Precision	Spiral Recall	Spiral F1-score	Elliptical Precision	Elliptical Recall	Elliptical F1-score
WISE images	WGC	89.03%	0.87	0.91	0.89	0.91	0.87	0.89
	ResNet50	87.38%	0.87	0.86	0.87	0.87	0.89	0.88
	VGG19	87.69%	0.87	0.87	0.87	0.88	0.88	0.87
	EfficientNet_b0	86.90%	0.85	0.89	0.87	0.89	0.85	0.87
	EfficientNet_b3	87.40%	0.87	0.87	0.87	0.88	0.88	0.88
	EfficientNet_v2	88.24%	0.87	0.88	0.88	0.89	0.88	0.89
	WGC_No_CBAM	88.48%	0.87	0.89	0.88	0.90	0.88	0.89
WISE magnitude	K-means	80.73%	0.90	0.71	0.79	0.74	0.91	0.82
	SVM	81.44%	0.76	0.91	0.82	0.89	0.73	0.80
WISE magnitude and images	WGC_mag	89.89%	0.87	0.92	0.90	0.93	0.88	0.90
SDSS images	SGC	94.64%	0.93	0.96	0.94	0.96	0.94	0.95
WISE images and SDSS images	MGMIC	95.39%	0.95	0.95	0.95	0.96	0.96	0.96

Figure 8(a) displays the galaxy clustering results obtained using the K-means algorithm, while Figure 8(b) showcases the decision boundaries defined by the optimized SVM model. Both K-means and SVM primarily classify galaxies based on the W2–W3 values. By employing the grid search optimization technique in Scikit-learn (Pedregosa et al. 2011), we identified the optimal parameter combination for the SVM model: a Linear kernel with a regularization parameter C set to 100. As indicated in Table 10, the accuracy of using the K-means algorithm along with W1–W2 and W2–W3 data is 8% lower than that achieved using the WGC network and images. Similarly, the accuracy of using SVM with W1–W2 and W2–W3 data is approximately 7% lower compared to the WGC network and images.

6. Analysis of Misclassified Samples

In this section the classification result is discussed. Since WGC and K-means use different data, these two methods are chosen to classify all data separately, and then the samples that could be correctly classified and incorrectly classified by both WGC and K-means are divided into four classes. The result is shown in Table 12. The WGC and K-means algorithms have the same classification criteria that the astronomical objects with few red substances around them are judged as spiral and those with a lot of red substances are elliptical. These red substances mainly originate from the W3 band that contains significant PAH features (Simonian & Martini 2017), which mean the elliptical cases are surrounded by large amounts of dust.

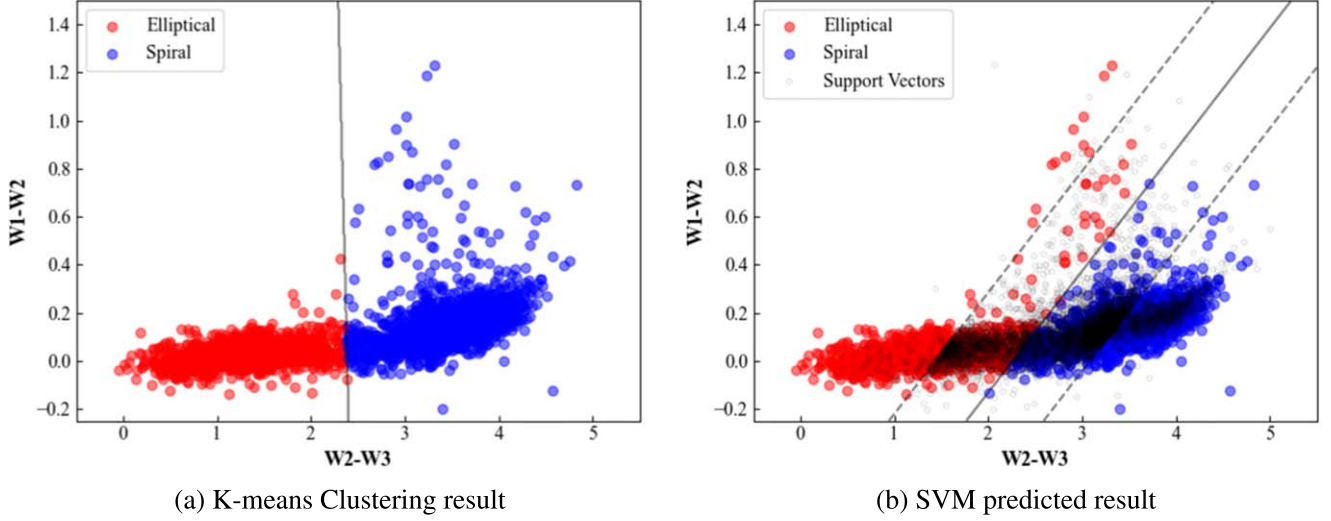


Figure 8. The figure illustrates the prediction outcomes using traditional machine learning methods, K-means and SVM. The decision boundary for K-means is delineated as a line almost perpendicular to the “W2–W3” axis, situated around $(W2-W3) = 2.4$, and is mathematically expressed as: $(W1-W2) = -0.0635 \times (W2-W3) + 2.3862$. For SVM, the decision boundary follows the formula: $(W1-W2) = 0.5055 \times (W2-W3) - 1.144$.

Table 11
Comparison of Experimental Results for Different Networks (2). Bold Entries in the Table Highlight the Best Results in each Column

Networks	Flops	Params	Training time	Total memory
WGC	284.66M	6.42M	2450.64s	5.05MB
ResNet50	338.09M	25.55M	4554.93s	8.96MB
VGG19	1.72G	143.67M	5044.53s	9.92MB
EfficientNet_b0	2.38M	1.32M	3054.31s	5.49MB
EfficientNet_b3	3.66M	1.63M	4300.82s	10.75MB
EfficientNet_v2	237.935M	20.180M	4195.96s	12.09M

In order to analyze the reason why the sample was misclassified and the common aspects between different types, the parameter H_α and spectral energy distribution (SED) are also cross-matched from the SDSS and WISE catalogs.

6.1. H_α EW

The H_α emission line of a galaxy can be used to trace star formation. The parameters H_{α_reqw} and H_{α_eqw} are cross-matched in the galSpecLine catalog of SDSS.⁴ REQW is the equivalent width of the continuum-subtracted emission line computed from straight integration over the bandpasses. The EQW measurement provides a better measure of the true equivalent width. The purpose of this measurement is to help characterize the stellar absorption that affects this line. At the same time, H_{α_EW} is also calculated which can characterize the stellar absorption affecting the line, and the formula is written in Equation (5). The distribution of H_{α_EW} is

shown in Figure 9

$$EW_{\text{stellar}} = REQW - EQW. \quad (5)$$

The H_{α_EW} distribution and images of ell_corr_ell and spir_err_ell are similar. The H_{α_EW} distributions of ell_err_spir and spir_corr_spir are similar. From the parameter H_{α_EW} , it is also difficult to classify ell_err_spir and spir_err_ell into correct classes. There is a certain relationship between the amount of dust and the distribution of H_{α_EW} . More dusty galaxies have a smaller value and smaller distribution intervals of H_{α_EW} than less dusty galaxies.

6.2. Spectral Energy Distribution (SED)

The magnitudes of five bands, u , g , r , i and z , are cross-matched in the SDSS catalog and the magnitudes of the seven bands J , h , k , $W1$, $W2$, $W3$ and $W4$ are cross-matched in the WISE catalog. The mean values of the 12 bands are also calculated. Figure 10 is a line chart of the mean of magnitude. The magnitudes of the four types of galaxies are similar in the visible wavelengths. In the near-infrared (NIR) and mid-

⁴ The SDSS parameters are available in casjob of SDSS, at <http://skyserver.sdss.org/CasJobs/SubmitJob.aspx>.

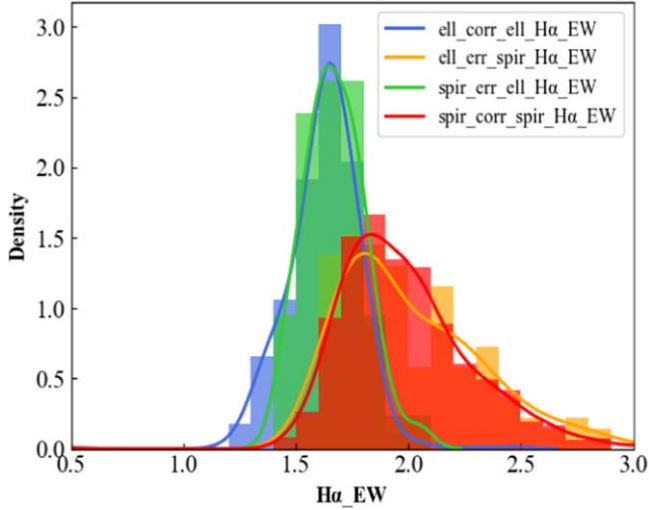


Figure 9. H_{α} -EW distributions of different types of galaxies. Due to the imbalance of the four types of sample data, the ordinate is set to density, which can compare the differences between the four types of samples more clearly.

infrared wavelengths, `ell_corr_ell` and `spir_err_ell` have a similar trend of change, and `spir_corr_spir` and `ell_err_spir` also have a similar trend of change. The SED curves of the four types of galaxies intersect at W2.

The above analysis shows that it is difficult to predict both `spir_err_ell` and `ell_err_spir` correctly as they are similar to the predicted categories in several ways. This could be due to a labeling error in Galaxy Zoo 2.

7. Analysis and Integration of Galaxy Morphology Classification Using WISE Infrared and SDSS Optical Images

This section delves deeper into the application of WGC in galaxy morphology classification. For this purpose, a new galaxy classifier based on the WGC structure, named the SGC, was trained using SDSS optical images corresponding to WGC training data. Figure 6 displays the accuracy changes in the SGC training and validation sets, showing a relatively stable rise in the accuracy curve of the SGC network. SGC demonstrates excellent classification performance, achieving an accuracy of 94.64% on the test set (see Table 10).

7.1. Analysis of the Consistency and Complementarity in Galaxy Morphology Classification between WISE Infrared Images and SDSS Optical Images

This study delves into the consistency and complementarity of WISE infrared images and SDSS optical images in galaxy morphology classification tasks. The prediction results of both classifiers on the test set are categorized into four scenarios: both WGC and SGC correctly predict (`WGC_corr&SGC_corr`), WGC correctly predicts while SGC errs (`WGC_corr&SGC_err`),

Table 12
Classification Result

Class	Number	Samples
<code>ell_corr_ell</code>	8468	
<code>ell_err_spir</code>	1441	
<code>spir_err_ell</code>	219	
<code>spir_corr_spir</code>	10313	

Note. The classification result is judged by the WGC and K-means algorithms. The galaxy is not counted if the results of the two algorithms are inconsistent. The parameter information is as follows. `ell_corr_ell`: Elliptical is judged as a Elliptical, `ell_err_spir`: Elliptical is judged as a Spiral, `spir_err_ell`: Spiral is judged as a Elliptical and `spir_corr_spir`: Spiral is judged as a Spiral.

WGC errs while SGC correctly predicts (`WGC_err&SGC_corr`), and both WGC and SGC err (`WGC_err&SGC_err`). The initial step involves the statistical analysis of prediction quantities (Table 13). When the predictions of both classifiers align (as shown on the diagonal of Table 13), the reliability of the predicted categories is exceptionally high, with an accuracy rate of 99.30%. The consistency between WGC and SGC classifiers can be leveraged to obtain more reliable category labels, thereby creating a data set with more accurate labels. This facilitates subsequent analyses of different types of galaxies.

Subsequently, a joint analysis of image features was conducted. For `WGC_corr&SGC_corr` cases (Table 14), it was observed that objects with more red matter in WISE images correspond to elliptical galaxies, while those with less red matter correspond to spiral galaxies. This finding aligns with the conclusions drawn from the Color-Color analysis in Section 5.

Special attention is given to samples where the two classifiers predict inconsistently. For `WGC_corr&SGC_err` cases (Table 14), this implies that WISE images might compensate for optical image classification errors. When the optical images of elliptical galaxies have brighter stars or other celestial bodies in the galaxy halo, it increases the likelihood of SGC mistaking elliptical galaxies for spirals. In such cases, WGC can correctly predict elliptical galaxies due to the evident red material in WISE images. When the spiral arms and halo of spiral galaxies in optical images are closely knit, it increases the likelihood of SGC mistaking spirals for ellipticals. At these times, the absence of diffuse red material or the presence of more green material in WISE images allows WGC to make correct predictions.

For the samples classified as `WGC_err&SGC_corr` (Table 14), in the case of elliptical galaxies, optical images show a smooth halo structure, while WISE images lack diffuse red material or have more green material. For spiral galaxies,

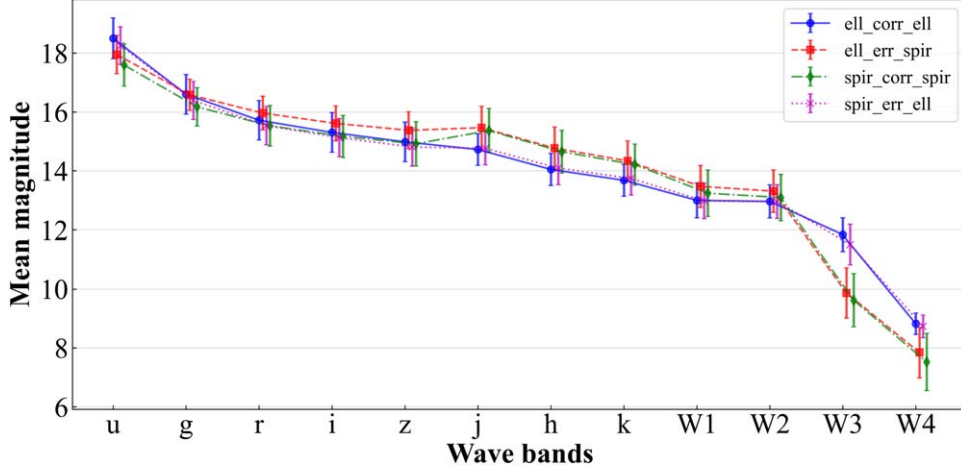


Figure 10. The bands u to z belong to SDSS, j to k belong to 2MASS and W1 to W4 belong to WISE. Order of wavelengths is from smallest to largest.

Table 13
Performance Comparison of WGC and SGC

	SGC_corr	SGC_err
WGC_corr	2125	121
WGC_err	275	15

Note. This table shows the comparison of classification results for elliptical and spiral galaxies using WGC and SGC. WGC_corr and WGC_err represent the samples of WGC's correct and incorrect predictions, respectively, similar to SGC_corr and SGC_err for SGC.

optical images have obvious spiral arm features, while WISE images contain diffuse red material.

Furthermore, for samples misclassified by both classifiers (WGC_err&SGC_err, Table 14), the specific features leading to errors differ between galaxy types. In the case of elliptical galaxies, misclassification is primarily due to non-smooth halo structures and, in WISE images, either the absence of diffuse red matter or the presence of diffuse green features. For spiral galaxies, the errors stem from the tightly bound spiral arms seen in SDSS images and the diffuse red matter observable in their WISE images.

In summary, the statistical consistency illustrates the role of combining WGC and SGC to obtain pure samples. At the same time, the variability in image features highlights the potential for complementarity between WGC and SGC in galaxy morphology classification.

7.2. The Multi-band Galaxy Morphology Integrated Classifier—MGMIC, Based on the Complementarity of WISE and SDSS Images

















































A stacking ensemble learning approach was employed to verify the complementarity of WGC and SGC in galaxy morphology classification tasks, resulting in the creation of the MGMIC. The

structure of this integrated learning classifier is illustrated in Figure 11. MGMIC first extracts flattened features from WGC and SGC and merges them through concatenation. To effectively handle these merged features, a stacking layer structure composed of three fully connected layers was designed, with the first two layers containing 2048 neurons each, followed by Dropout and ReLU activation functions. The last fully connected layer, serving as the output layer, contains two neurons. This stacking layer is trained using a cross-entropy loss function and an Adam optimizer. Figure 6 shows the accuracy changes of the MGMIC training and validation sets. MGMIC converged after a few epochs of slight increase and overall showed no significant fluctuations. The performance evaluation on the test data set indicates that the accuracy of MGMIC reaches 95.39%, which is 0.75% higher than using SGC alone. Moreover, it surpasses SGC on most metrics (as seen in Table 10), effectively integrating the inconsistent predictions of SGC and WGC. These results confirm the complementarity of WISE infrared images and SDSS optical images in galaxy morphology classification tasks.

8. HyperLEDA Data Set

To explore the effect of the WGC network on other data sets and which types of spiral galaxies are classified with elliptical galaxies when the WGC network fails, this section uses the WGC network to classify the Hyper-Linked Extragalactic Databases and Archives (HyperLEDA) data set (Paturel et al. 2003). The HyperLEDA data set contains millions of galaxies and morphological information, which has more explicit morphological information with a longer redshift range compared to Galaxy Zoo 2. In this experiment, the maximum redshift is 0.6. The HyperLEDA data set classifies galaxies into 12 categories: E, E-S0, S0, S0-a, Sa, Sab, Sb, Sbc, Sc, Scd, Sd and Sm. In this section, the WGC network is used to perform automated morphological classification of different types of galaxies.

Table 14
Classification Results of WGC and SGC

Class	Spiral						Elliptical					
SGC_corr&WGC_corr												
SGC_corr&WGC_err												
SGC_err&WGC_corr												
SGC_err&WGC_err												

Note. This table shows the comparison of classification results for elliptical and spiral galaxies using WGC and SGC. The four rows represent the scenarios where both classifiers predict correctly, SGC predicts correctly while WGC errs, SGC predicts incorrectly while WGC is correct, and both classifiers predict incorrectly, from top to bottom respectively. The two columns display spiral galaxies and elliptical galaxies. Each group contains three pairs of sample images, with each pair comprising corresponding SDSS and WISE images.

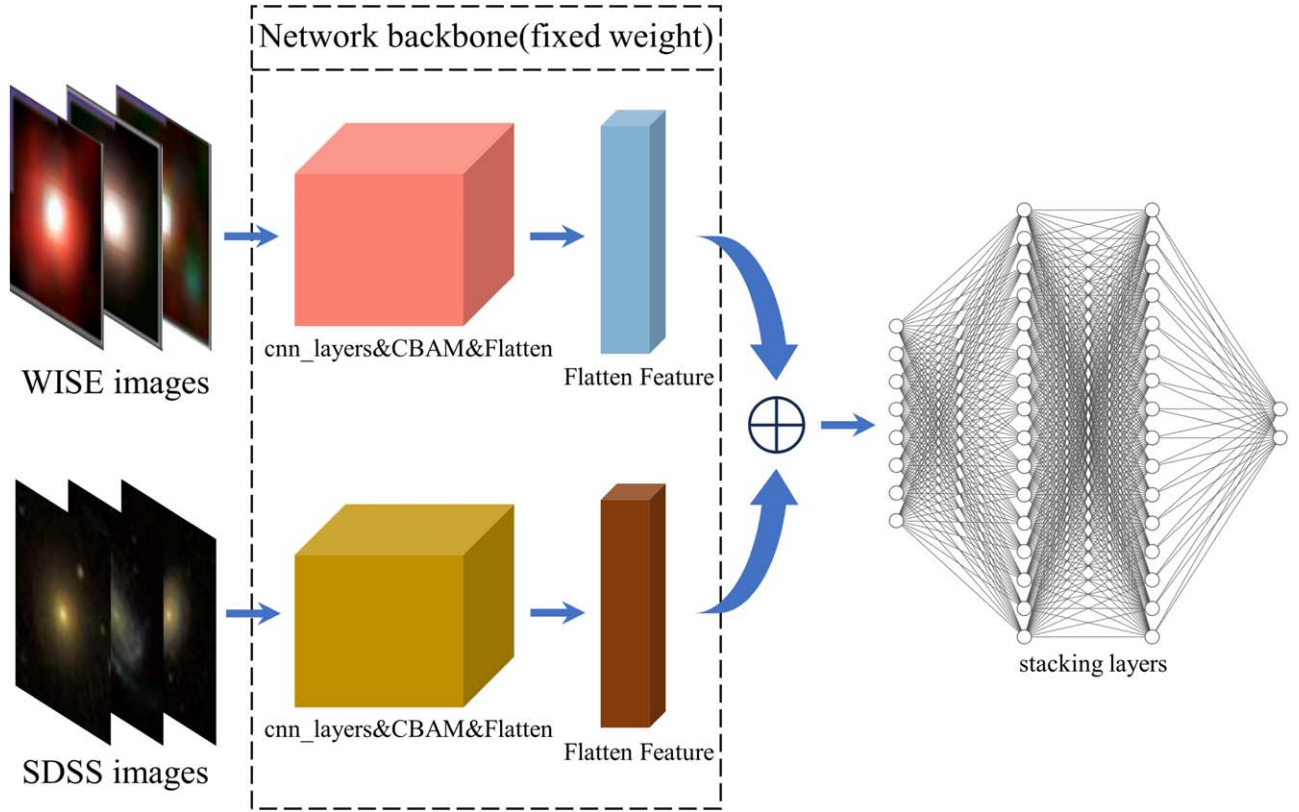


Figure 11. Structure Diagram of the MGMIC. The MGMIC consists of two components: the pre-trained WGC and SGC, and the stacking layer that is yet to be trained. Initially, WGC and SGC independently process WISE and SDSS images to obtain flattened features. These concatenated flattened features are then used to train the fully connected layers within the stacking layers.

8.1. Introduction to the Data Set

In this experiment, both classes E and E-S0 are categorized as elliptical galaxies, and the different classes of spiral galaxies are classified separately from elliptical

galaxies in ten comparison experiments. In addition, due to the small number of some classes of spiral galaxies, the data set division for this experiment is 8:1:1. Detailed information is given in Table 15.

Table 15
Dataset Segmentation

Class	Training Set	Validation Set	Test Set
E	7681	959	960
S0	7549	943	943
S0-a	7945	993	993
Sa	3997	499	499
Sab	8409	1051	1051
Sb	7988	998	998
Sbc	7969	995	996
Sc	7932	991	991
Scd	4457	557	557
Sd	3079	384	385
Sm	2317	289	290

Table 16
Comparison of Experimental Results with Different Dataset

Dataset	Accuracy	Spiral Precision	Spiral Recall	Spiral F1- score	Elliptical Precision	Elliptical Recall	Elliptical F1-score
E-S0	68.94%	0.68	0.69	0.69	0.69	0.69	0.69
E-S0_a	78.39%	0.76	0.84	0.80	0.81	0.73	0.77
E-Sa	79.36%	0.75	0.60	0.67	0.81	0.89	0.85
E-Sab	76.08%	0.76	0.80	0.78	0.77	0.72	0.74
E-Sb	85.18%	0.85	0.86	0.86	0.85	0.84	0.85
E-Sbc	86.40%	0.85	0.84	0.86	0.85	0.88	0.86
E-Sc	90.41%	0.90	0.91	0.91	0.91	0.89	0.90
E-Scd	90.04%	0.86	0.87	0.87	0.93	0.92	0.92
E-Sd	90.26%	0.83	0.83	0.83	0.93	0.93	0.93
E-Sm	87.76%	0.76	0.69	0.72	0.91	0.93	0.92

8.2. Analysis of Classification Results

Table 16 shows the results of the ten experiments, and this subsection focuses on the precision of the two types of galaxies due to the relatively small number of some types of spiral galaxies in the database, which leads to unevenness in the whole training sample. From Table 16, it can be seen that the classification precisions of elliptical galaxies and spiral galaxies in Sc, Sbc, Sb, Scd and Sd are more ideal, which can be comparable to the accuracy rate of Galaxy Zoo 2. It can also be shown that these types of spiral galaxies have relatively large morphological differences from elliptical galaxies in the infrared band, and can be automatically classified with neural networks.

9. Conclusion

This paper introduces the WGC for classifying spiral and elliptical galaxies. A total of 12,088 spiral galaxies and 12,852 elliptical galaxies were selected, and the data set was divided into training, validation and test sets in a 7:2:1 ratio. The WGC network achieved a classification accuracy of 89.03% on the test set, with accuracies of 87% for spiral galaxies and 91% for

elliptical galaxies. Ablation experiments demonstrated the effectiveness of model improvement modules, and comparisons with other classification networks confirmed WGC's superior performance. Adopting the principle of multimodal feature integration, the WGC_mag network was designed, integrating WISE image features with magnitude characteristics to achieve a high accuracy rate of 89.89%.

For the Color-Color diagram, SVM and K-means were used to find the optimal classification boundary, with the high-performing SVM achieving an accuracy of 81.44%. This indicates that galaxy morphology classification accuracy using magnitude information is lower than WGC, demonstrating the benefit of using WISE images for WISE galaxy classification.

The article explores the criteria WGC uses to distinguish galaxy morphologies. It was observed that WGC classifies images with more dust as elliptical galaxies, corresponding to lower H_{α_EW} and aligning with elliptical galaxies' typically lower star formation rates. Conversely, images with less diffuse dust are classified as spiral galaxies by WGC, corresponding to higher H_{α_EW} values, in line with the higher star formation rates of spiral galaxies. The high overlap in SED and H_{α_EW} between misclassified samples and confused categories

demonstrates that WGC has learned to distinguish galaxy morphologies in WISE images as effectively as possible.

The paper further investigates the consistency and complementarity of WISE infrared images and SDSS optical images in the task of galaxy morphology classification. Within the WGC framework, an optical image galaxy morphology classifier, SGC, was trained using the corresponding SDSS images from the WISE data set, achieving an accuracy of 94.64%. For predictions where SGC and WGC concur, the network's accuracy reaches an impressive 99.30%, making it suitable for obtaining a clean galaxy data set. Capitalizing on the complementary features of WISE infrared images and SDSS optical images, a stacking ensemble approach was utilized to integrate the WGC and SGC classifiers, resulting in an ensemble learning classifier—MGMIC—with an accuracy of 95.39%.

Finally, the versatility of WGC in other data sets was verified. On the HyperLEDA data set, the distinction between elliptical galaxies and Sc, Scd and Sd spiral galaxies was most pronounced, with an accuracy reaching 90%, outperforming the classification results of the Galaxy Zoo 2 labeled WISE data set.

In this paper, we only use three bands for image synthesis classification. Each band contains different information, and whether they can be distinguished when adding the W4 band and how to use such information are also our future works. Furthermore, expanding the classification to encompass more galaxy morphological categories using WISE images is a focus for our future work.

Acknowledgments

This work is supported by the Joint Research Fund in Astronomy, National Natural Science Foundation of China (NSFC, grant No. U1931134), the Natural Science Foundation of Hebei, A2020202001, and the Natural Science Foundation of Tianjin Municipality, 22JCYBJC00410.

References

Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, *ApJS*, **249**, 3

- Buades, A., Coll, B., & Morel, J.-M. 2005, in 2005 IEEE Comp. Soc. Conf. Computer Vision and Pattern Recognition (CVPR'05), 2 (San Diego, CA: IEEE), 60
- Burke, C. J., Aleo, P. D., Chen, Y.-C., et al. 2019, *MNRAS*, **490**, 3952
- Curran, S. J. 2020, *MNRAS*, **493**, L70
- Dalal, N., & Triggs, B. 2005, in 2005 IEEE Comp. Soc. Conf. Computer Vision and Pattern Recognition (CVPR'05), 1 (San Diego, CA: IEEE), 886
- Dieleman, S., Willett, K. W., & Dambre, J. 2015, *MNRAS*, **450**, 1441
- Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., Tuccillo, D., & Fischer, J. L. 2018, *MNRAS*, **476**, 3661
- Faisst, A. L., Prakash, A., Capak, P. L., & Lee, B. 2019, *ApJL*, **881**, L9
- Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, *SSRv*, **123**, 485
- Glorot, X., Bordes, A., & Bengio, Y. 2011, in Proc. 14th Int. Conf. Artificial Intelligence and Statistics, 15 (Fort Lauderdale, FL: PMLR), 315
- González, R. E., Muñoz, R. P., & Hernández, C. A. 2018, *A&C*, **25**, 103
- Gupta, R., Srijith, P. K., & Desai, S. 2022, *A&C*, **38**, 100543
- He, K., Zhang, X., Ren, S., & Sun, J. 2015, in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (Las Vegas, NV: IEEE), 770
- He, Z., Qiu, B., Luo, A. L., et al. 2021, *MNRAS*, **508**, 2039
- Kalvankar, S., Pandit, H., & Parwate, P. 2020, arXiv:2008.13611
- Kunsági-Máté, S., Beck, R., Szapudi, I., & Csabai, I. 2022, *MNRAS*, **516**, 2662
- Lowe, D. G. 2004, *Int. J. Comput. Vision*, **60**, 91
- Lupton, R., Blanton, M. R., Fekete, G., et al. 2004, *PASP*, **116**, 133
- Kessler, M. F., Metcalfe, L., & Salama, A. 1992, *Space Sci Rev*, **61**, 45–60
- Mittal, A., Soorya, A., Nagrath, P., & Hemanth, D. J. 2020, *EScIn*, **13**, 601
- Neugebauer, G., Habing, H. J., van Duinen, R., et al. 1984, *ApJL*, **278**, L1
- Ojala, T., Pietikainen, M., & Maenpää, T. 2002, *ITPAM*, **24**, 971
- Paturel, G., Petit, C., Prugniel, P., et al. 2003, *A&A*, **412**, 45
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *Journal of Machine Learning Research*, **12**, 2825
- Shi, J.-H., Qiu, B., Luo, A. L., et al. 2022, *MNRAS*, **516**, 264
- Simonian, G. V., & Martini, P. 2017, *MNRAS*, **464**, 3920
- Simonyan, K., & Zisserman, A. 2014, arXiv:1409.1556
- Tan, M., & Le, Q. V. 2019, in Proc. 36th Int. Conf. Mach. Learn., 97 (Long Beach, CA: PMLR), 6105
- Tan, M., & Le, Q. V. 2021, in Proc. 38th International Conf. Mach. Learn., 139, (PMLR), 10096
- Taylor, M. B. 2005, in ASP Conf. Ser. 347, *Astronomical Data Analysis Software and Systems XIV*, ed. P. Shopbell, M. Britton, & R. Ebert (San Francisco, CA: ASP), 29
- Viola, P., & Jones, M. 2001, in Proc. 2001 IEEE Comp. Soc. Conf. Computer Vision and Pattern Recognition, 1 (Kauai, HI: IEEE)
- Werner, M. W., Roellig, T. L., Low, F. J., et al. 2004, *ApJS*, **154**, 1
- Willett, K. W., Lintott, C. J., Bamford, S. P., et al. 2013, *MNRAS*, **435**, 2835
- Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. 2018, in Proc. Eur. Conf. Computer Vision (ECCV) (Berlin: Springer), 3
- Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, *AJ*, **140**, 1868
- Zhang, Z., Zou, Z., Li, N., & Chen, Y. 2022, *RAA*, **22**, 055002
- Zhao, G., Qiu, B., Luo, A. L., et al. 2023, *RAA*, **23**, 085010
- Zhu, X.-P., Dai, J.-M., Bian, C.-J., et al. 2019, *Astrophysics and Space Science*, **364**, 55